

0590  
1120

# 9



OIPE

## RAW SEQUENCE LISTING

DATE: 11/29/2002

PATENT APPLICATION: US/09/832,464

TIME: 16:03:01

Input Set : N:\Crf3\RULE60\09832464.raw

Output Set: N:\CRF4\11292002\I832464.raw

## SEQUENCE LISTING

## 3 (1) GENERAL INFORMATION:

5 (i) APPLICANT: de Boer, Piet A.J.

6 Hale, Cynthia A.

8 (ii) TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR SCREENING

9 ANTIMICROBIALS

11 (iii) NUMBER OF SEQUENCES: 25

13 (iv) CORRESPONDENCE ADDRESS:

14 (A) ADDRESSEE: MEDLEN &amp; CARROLL

15 (B) STREET: 220 Montgomery Street, Suite 2200

16 (C) CITY: San Francisco

17 (D) STATE: California

18 (E) COUNTRY: United States of America

19 (F) ZIP: 94104

21 (v) COMPUTER READABLE FORM:

22 (A) MEDIUM TYPE: Floppy disk

23 (B) COMPUTER: IBM PC compatible

24 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

25 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

27 (vi) CURRENT APPLICATION DATA:

C--&gt; 28 (A) APPLICATION NUMBER: US/09/832,464

C--&gt; 29 (B) FILING DATE: 11-Apr-2001

30 (C) CLASSIFICATION:

32 (vii) PRIOR APPLICATION DATA:

33 (A) APPLICATION NUMBER: US/09/184,826

34 (B) FILING DATE:

36 (viii) ATTORNEY/AGENT INFORMATION:

37 (A) NAME: Carroll, Peter G.

38 (B) REGISTRATION NUMBER: 32,837

39 (C) REFERENCE/DOCKET NUMBER: CASE-02249

41 (ix) TELECOMMUNICATION INFORMATION:

42 (A) TELEPHONE: (415) 705-8410

43 (B) TELEFAX: (415) 397-8338

46 (2) INFORMATION FOR SEQ ID NO: 1:

48 (i) SEQUENCE CHARACTERISTICS:

49 (A) LENGTH: 2160 base pairs

50 (B) TYPE: nucleic acid

51 (C) STRANDEDNESS: single

52 (D) TOPOLOGY: linear

54 (ii) MOLECULE TYPE: DNA (genomic)

59 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

61 CAATACCAGG GATGAAGTAA AGAATTAGTA ATACAATTGC GCGCGGCAGA TACCAGGCAA 60

63 ATTTTGGCCA TTCGCGTTTC ATGATTCGCG GCACATCTTT CATGATACCG AAAATCCCGG 120

ENTERED

## RAW SEQUENCE LISTING

DATE: 11/29/2002

PATENT APPLICATION: US/09/832,464

TIME: 16:03:01

Input Set : N:\Crif3\RULE60\09832464.raw

Output Set: N:\CRF4\11292002\I832464.raw

```

65 TATCTGGCGG TGTAGCGCCA GTCAATCGTG CTTCCAGTTG TTCAGCCAAT AAACCGTTAA 180
67 ACGGAGCGGC AATCCAGTTA GCAATCGTGG AGAAGAAATA GCCAAACACT AACAGCACAG 240
69 AGATGACACG CAGAGGCCAC AACAGATAAC TCAGCCATTG TAGCCAGTCC GGAACGTAAC 300
71 TCATGAGAGT CGGGATCCAG ACATCGAGCT GTGTAAAGAG CCACCAGAAT GCGCCCCCA 360
73 TCAACAAAAT ATTGACCAGC AGCGGTAAAA TAACGAAACG CCGAATCCCA GGTTGCGAGA 420
75 CGAGCTTCCA GCCTTGCGCA AAATAGTAAA AACCCTGCGT TGGGGCAGAT GTGAATGATG 480
77 AAACCATAAT CAGGATGAGC TCCTTTTGAC CAATCCCAGG AAAATTCTGC GTATTTTACC 540
79 GGGTAATTGC GCAATGGACA GTTAGGATAT GTTCGAAAAA ACAGCAAAA GCACGATTTC 600
81 ATCTATCTTT GTGCTGTGAA AGTTAATAGT GCACTTGAC TTAGGTAAT CCGCAAATAC 660
83 TCTTAGTGAG TAAATGTTTG CCGTGGTGGC AAGGTGTTAG AACAACAGAG AATATAATGA 720
85 TGCAGGATTT GCGTCTGATA TTAATCATTT TTGGCGCGAT CGCCATAATC GCTTTACTGG 780
87 TACATGGTTT CTGGACCAGC CGTAAAGAAC GATCTTCTAT GTTCCGCGAT CGGCCATTAA 840
89 AACGAATGAA GTCAAAACGT GACGACGATT CTTATGACGA GGATGTCGAA GATGATGAGG 900
91 GCGTTGGTGA GGTTCTGTGT CACCGCGTGA ATCATGCCCC GGCTAACGCT CAGGAGCATG 960
93 AGGCTGCTCG TCCGTCGCCG CAACACCACT ACCAACCGCC TTATGCGTCT GCGCAGCCGC 1020
95 GTCAACCGGT CCAGCAGCCG CCTGAAGCGC AGGTACCGCC GCAACATGCT CCGCATCCAG 1080
97 CGCAGCCGGT GCAGCAGCCT GCCTATCAGC CGCAGCCTGA ACAGCCGTTG CAGCAGCCAG 1140
99 TTTCCGCCACA GGTCGCGCCA GCGCCGCGC CTGTGCATTG AGCACCGCAA CCGGCACAAC 1200
101 AGGCTTTCCA GCCTGCAGAA CCCGTAGCGG CACCACAGCC TGAGCCTGTA GCGGAACCTG 1260
103 CTCCAGTTAT GGATAAACC GAGCGCAAAG AAGCGGTGAT TATCATGAAC GTCGCGGCGC 1320
105 ATCACGGTAG CGAGCTAAAC GGTGAAGCTC TTCTTAACAG CATTCAACAA GCGGGCTTCA 1380
107 TTTTGGCGCA TATGAATATT TACCATCGTC ATCTTAGCCC GGATGGCAGC GGCCCGGCGT 1440
109 TATTCAGCCT GGCGAATATG GTGAAACCGG GAACCTTTGA TCCTGAAATG AAGGATTTC 1500
111 CTACTCCGGG TGTCATATC TTTATGCAGG TACCGTCTTA CCGTGACGAG CTGCAGAACT 1560
113 TCAAGCTGAT GCTGCAATCT GCGCAGCATA TTGCCGATGA AGTGGGCGGT GTCGTGCTTG 1620
115 ACGATCAGCG CCGTATGATG ACTCCGCAGA AATTGCGCGA GTACCAGGAC ATCATCCGCG 1680
117 AAGTCAAAGA CGCCAACGCC TGATACACTT AAGGCAAATT AACTCCTCTT CGAACCCCGG 1740
119 CTTGTGCGGG GTTTTGTAGCA TTGATGGTGC GATATGGAAT CAATCGAACA ACAACTGACA 1800
121 GAACTGCGAA CGACGCTTCG CCATCATGAA TATCTTTATC ATGTGATGGA TGCGCCGGAA 1860
123 ATTCCCGACG CTGAATACGA CAGGCTGATG CGCGAACTGC GCGAGCTGGA AACCACACAT 1920
125 CCAGAACTGA TTACGCCTGA TTCGCCTACT CAACGTGTAG GCGCTGCGCC GCTGGCGGCT 1980
127 TTCAGCCAGA TACGCCATGA AGTACCAATG CTGTCACTGG ATAACGTTTT TGATGAAGAA 2040
129 AGCTTTCTTG CTTTCAACAA ACGTGTGCAG GACCGTCTGA AAAACAACGA GAAAGTCACC 2100
131 TGGTGCTGTG AGCTGAAGCT GGATGGTCTT GCCGTCAGTA TTCTGTATGA AAATGGCGTT 2160

```

134 (2) INFORMATION FOR SEQ ID NO: 2:

136 (i) SEQUENCE CHARACTERISTICS:

137 (A) LENGTH: 328 amino acids

138 (B) TYPE: amino acid

139 (C) STRANDEDNESS: Not Relevant

W--&gt; 140 (D) TOPOLOGY: Not Relevant

142 (ii) MOLECULE TYPE: protein

147 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

149 Met Met Gln Asp Leu Arg Leu Ile Leu Ile Val Gly Ala Ile Ala

150 1 5 10 15

152 Ile Ile Ala Leu Leu Val His Gly Phe Trp Thr Ser Arg Lys Glu Arg

153 20 25 30

155 Ser Ser Met Phe Arg Asp Arg Pro Leu Lys Arg Met Lys Ser Lys Arg

156 35 40 45

158 Asp Asp Asp Ser Tyr Asp Glu Asp Val Glu Asp Asp Glu Gly Val Gly

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/832,464

DATE: 11/29/2002

TIME: 16:03:01

Input Set : N:\Crif3\RULE60\09832464.raw

Output Set: N:\CRF4\11292002\I832464.raw

```

159          50          55          60
161  Glu Val Arg Val His Arg Val Asn His Ala Pro Ala Asn Ala Gln Glu
162  65          70          75          80
164  His Glu Ala Ala Arg Pro Ser Pro Gln His Gln Tyr Gln Pro Pro Tyr
165          85          90          95
167  Ala Ser Ala Gln Pro Arg Gln Pro Val Gln Gln Pro Pro Glu Ala Gln
168          100         105         110
170  Val Pro Pro Gln His Ala Pro His Pro Ala Gln Pro Val Gln Gln Pro
171          115         120         125
173  Ala Tyr Gln Pro Gln Pro Glu Gln Pro Leu Gln Gln Pro Val Ser Pro
174          130         135         140
176  Gln Val Ala Pro Ala Pro Gln Pro Val His Ser Ala Pro Gln Pro Ala
177          145         150         155         160
179  Gln Gln Ala Phe Gln Pro Ala Glu Pro Val Ala Ala Pro Gln Pro Glu
180          165         170         175
182  Pro Val Ala Glu Pro Ala Pro Val Met Asp Lys Pro Lys Arg Lys Glu
183          180         185         190
185  Ala Val Ile Ile Met Asn Val Ala Ala His His Gly Ser Glu Leu Asn
186          195         200         205
188  Gly Glu Ala Leu Leu Asn Ser Ile Gln Gln Ala Gly Phe Ile Phe Gly
189          210         215         220
191  Asp Met Asn Ile Tyr His Arg His Leu Ser Pro Asp Gly Ser Gly Pro
192          225         230         235         240
194  Ala Leu Phe Ser Leu Ala Asn Met Val Lys Pro Gly Thr Phe Asp Pro
195          245         250         255
197  Glu Met Lys Asp Phe Thr Thr Pro Gly Val Thr Ile Phe Met Gln Val
198          260         265         270
200  Pro Ser Tyr Gly Asp Glu Leu Gln Asn Phe Lys Leu Met Leu Gln Ser
201          275         280         285
203  Ala Gln His Ile Ala Asp Glu Val Gly Gly Val Val Leu Asp Asp Gln
204          290         295         300
206  Arg Arg Met Met Thr Pro Gln Lys Leu Arg Glu Tyr Gln Asp Ile Ile
207          305         310         315         320
209  Arg Glu Val Lys Asp Ala Asn Ala
210          325

```

212 (2) INFORMATION FOR SEQ ID NO: 3:

214 (i) SEQUENCE CHARACTERISTICS:

215 (A) LENGTH: 327 amino acids

216 (B) TYPE: amino acid

217 (C) STRANDEDNESS: Not Relevant

W--&gt; 218 (D) TOPOLOGY: Not Relevant

220 (ii) MOLECULE TYPE: protein

225 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```

227  Asp Leu Asn Thr Ile Leu Ile Ile Val Gly Ile Val Ala Leu Val Ala
228  1          5          10          15
230  Leu Ile Val His Gly Leu Trp Ser Asn Arg Arg Glu Lys Ser Lys Tyr
231          20          25          30
233  Phe Asp Lys Ala Asn Lys Phe Asp Arg Thr Ser Leu Thr Ser Arg Ser
234          35          40          45

```

## RAW SEQUENCE LISTING

DATE: 11/29/2002

PATENT APPLICATION: US/09/832,464

TIME: 16:03:01

Input Set : N:\Crif3\RULE60\09832464.raw

Output Set: N:\CRF4\11292002\I832464.raw

```

236   His Thr Gln Glu Glu Met Val Gln Pro Asn Asn Ile Ser Pro Asn Thr
237       50                      55                      60
239   Tyr Val Glu Asn Gly His Thr Pro Ile Pro Gln Pro Thr Thr Glu Lys
240       65                      70                      75                      80
242   Leu Pro Ser Glu Ala Glu Leu Ile Asp Tyr Arg Gln Ser Asp Lys Ser
243               85                      90                      95
245   Val Asp Asp Ile Lys Ile Ser Ile Pro Asn Thr Gln Pro Ile Tyr Asp
246               100                    105                    110
248   Met Gly Asn His Arg Ser Glu Pro Ile Gln Pro Thr Gln Pro Gln Tyr
249               115                    120                    125
251   Asp Met Pro Thr Ala Asn Asn Val Ala Ser Met Thr Leu Glu Gln Leu
252       130                    135                    140
254   Glu Ala Gln Ser Gln Asn Val Gly Phe Asn Gly Ile Asn Ser Ser Ser
255       145                    150                    155                    160
257   Pro Glu Leu Arg Val Gln Leu Ala Glu Leu Ser His Glu Glu His Gln
258               165                    170                    175
260   Val Asp Tyr Asn Leu Ser Phe Asn Glu Pro Lys Ala Glu Thr Thr Ala
261               180                    185                    190
263   His Pro Lys Gln Thr Thr Gly Tyr Ile Gln Leu Tyr Leu Ile Pro Lys
264               195                    200                    205
266   Ser Ser Glu Glu Phe Asn Gly Ala Lys Leu Val Gln Ala Leu Glu Asn
267       210                    215                    220
269   Leu Gly Phe Ile Leu Gly Lys Asp Glu Met Tyr His Arg His Leu Asp
270       225                    230                    235                    240
272   Leu Ser Val Ala Ser Pro Val Leu Phe Ser Val Ala Asn Leu Glu Gln
273               245                    250                    255
275   Pro Gly Thr Phe Asn Ala Tyr Asn Leu Ala Glu Phe Asn Thr Ile Gly
276               260                    265                    270
278   Ile Val Leu Phe Met Gln Leu Pro Ser Pro Gly Asn Asn Leu Ala Asn
279               275                    280                    285
281   Leu Arg Met Met Met Arg Ala Ala His Thr Leu Ala Glu Asp Leu Gln
282       290                    295                    300
284   Gly Val Ile Leu Thr Glu Glu Gln Glu Ile Phe Asp Ala Asn Ala Glu
285       305                    310                    315                    320
287   Gln Ala Tyr Leu Ala Arg Val
288               325

```

290 (2) INFORMATION FOR SEQ ID NO: 4:

292 (i) SEQUENCE CHARACTERISTICS:

293 (A) LENGTH: 6 amino acids

294 (B) TYPE: amino acid

295 (C) STRANDEDNESS: Not Relevant

W--&gt; 296 (D) TOPOLOGY: Not Relevant

298 (ii) MOLECULE TYPE: peptide

303 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

305 Ile Leu Ile Ile Val Gly

306 1 5

308 (2) INFORMATION FOR SEQ ID NO: 5:

310 (i) SEQUENCE CHARACTERISTICS:

311 (A) LENGTH: 10 amino acids

## RAW SEQUENCE LISTING

DATE: 11/29/2002

PATENT APPLICATION: US/09/832,464

TIME: 16:03:01

Input Set : N:\CrF3\RULE60\09832464.raw

Output Set: N:\CRF4\11292002\I832464.raw

```

312         (B) TYPE: amino acid
313         (C) STRANDEDNESS: Not Relevant
W--> 314         (D) TOPOLOGY: Not Relevant
316         (ii) MOLECULE TYPE: peptide
319         (ix) FEATURE:
320             (A) NAME/KEY: Modified-site
321             (B) LOCATION: 3
322             (D) OTHER INFORMATION: /note= "The peptide at this
323 location can be either Arg or Asn."
325         (ix) FEATURE:
326             (A) NAME/KEY: Modified-site
327             (B) LOCATION: 4
328             (D) OTHER INFORMATION: /note= "The peptide at this
329 location can be either Leu or Thr."
332         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
W--> 334         Asp Leu Xaa Xaa Ile Leu Ile Ile Val Gly
335             1             5             10
337 (2) INFORMATION FOR SEQ ID NO: 6:
339         (i) SEQUENCE CHARACTERISTICS:
340             (A) LENGTH: 13 amino acids
341             (B) TYPE: amino acid
342             (C) STRANDEDNESS: Not Relevant
W--> 343         (D) TOPOLOGY: Not Relevant
345         (ii) MOLECULE TYPE: peptide
348         (ix) FEATURE:
349             (A) NAME/KEY: Modified-site
350             (B) LOCATION: 7
351             (D) OTHER INFORMATION: /note= "The peptide at this
352 location can be either"
354         (ix) FEATURE:
355             (A) NAME/KEY: Modified-site
356             (B) LOCATION: 8
357             (D) OTHER INFORMATION: /note= "The peptide at this
358 location can be either Ile or Val."
360         (ix) FEATURE:
361             (A) NAME/KEY: Modified-site
362             (B) LOCATION: 10
363             (D) OTHER INFORMATION: /note= "The peptide at this
364 locaiton can be either Ile or Leu."
366         (ix) FEATURE:
367             (A) NAME/KEY: Modified-site
368             (B) LOCATION: 11
369             (D) OTHER INFORMATION: /note= "The peptide at this
370 location can be either Ile or Val."
373         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
W--> 375         Ile Leu Ile Ile Val Gly Xaa Xaa Ala Xaa Xaa Ala Leu
376             1             5             10
378 (2) INFORMATION FOR SEQ ID NO: 7:
380         (i) SEQUENCE CHARACTERISTICS:

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/09/832,464

DATE: 11/29/2002  
TIME: 16:03:02

Input Set : N:\Crf3\RULE60\09832464.raw  
Output Set: N:\CRF4\11292002\I832464.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:5; Xaa Pos.3,4  
Seq#:6; Xaa Pos.7,8,10,11  
Seq#:7; Xaa Pos.7,8,10,11,14  
Seq#:8; Xaa Pos.3  
Seq#:9; Xaa Pos.3,7  
Seq#:11; Xaa Pos.2,6  
Seq#:13; Xaa Pos.1  
Seq#:14; Xaa Pos.1,5

## VERIFICATION SUMMARY

PATENT APPLICATION: US/09/832,464

DATE: 11/29/2002

TIME: 16:03:02

Input Set : N:\Crf3\RULE60\09832464.raw

Output Set: N:\CRF4\11292002\I832464.raw

L:28 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]  
L:29 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]  
L:140 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=2  
L:218 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=3  
L:296 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=4  
L:314 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=5  
L:334 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5 after pos.:0  
L:343 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=6  
L:375 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6 after pos.:0  
L:384 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=7  
L:422 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7 after pos.:0  
L:434 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=8  
L:448 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8 after pos.:0  
L:457 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=9  
L:477 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9 after pos.:0  
L:486 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=10  
L:504 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=11  
L:524 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11 after pos.:0  
L:533 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=12  
L:551 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=13  
L:565 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:13 after pos.:0  
L:574 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=14  
L:594 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:14 after pos.:0  
L:654 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=18  
L:672 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=19  
L:699 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=20  
L:717 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=21  
L:738 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=22  
L:759 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=23